CSE 392/CS 395T/M 397C: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin

Spring 2025

Supplement: Spectral sums

Outline

Spectral Sums

Given a symmetric positive semidefinite (PSD) matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ with eigen-decompostion $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$ and eigenvalues $\{\lambda_i\}_{i=1}^d$, and desired function $f(\cdot)$, compute the *trace of the matrix function* $f(\boldsymbol{A}) = \boldsymbol{U} f(\Lambda)\boldsymbol{U}^\top$, i.e.,

$$\mathrm{Tr}(f(\boldsymbol{A})) = \sum_{i=1}^{d} f(\lambda_i).$$

- *Popular examples:* log-determinant ($\log(x)$), numerical rank (step function), spectral density $\delta(x - \lambda_i)$, Schatten p-norms ($x^{p/2}$), von Neumann Entropy ($x\log(x)$), Estrada index ($\exp(x)$), trace of matrix inverse ($\frac{1}{x}$).
- *Applications:* machine learning, graph signal processing, quantum algorithms, scientific computing, statistics, computational biology and physics.
- *Naive approaches :* Eigenvalue decomposition, Cholesky Decomposition, singular value decomposition (SVD).
  Cost: $O(d^3)$ or [Theory: $O(d^\omega)$ and $\omega = 2.373$].

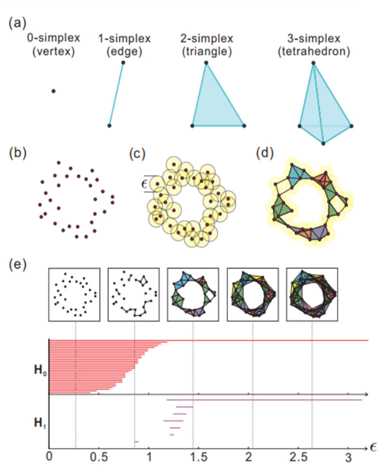
Spectral Sums

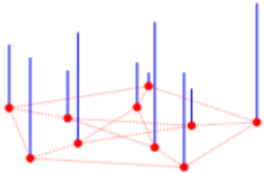
PSD matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ with eigen-decomposition $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^T$ and eigenvalues $\{\lambda_i\}_{i=1}^d$, and desired function $f(\cdot)$, compute the *trace of the matrix function* $f(\boldsymbol{A}) = \boldsymbol{U}f(\Lambda)\boldsymbol{U}^\top$, i.e.,

$$\mathrm{Tr}(f(\boldsymbol{A})) = \sum_{i=1}^{d} f(\lambda_i).$$

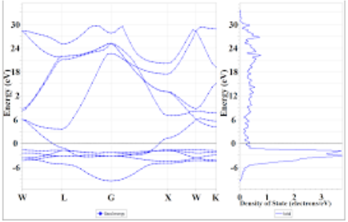
- *Popular examples:* log-determinant ($\log(x)$), numerical rank (step function), spectral density $\delta(x-\lambda_i)$, Schatten p-norms ($x^{p/2}$), von Neumann Entropy ($x\log(x)$), Estrada index ($\exp(x)$), trace of matrix inverse ($\frac{1}{x}$).
- Approximate the function using two approaches:
  1. Chebyshev polynomial approximation
  2. Lanczos quadrature method

Spectral Sums Applications

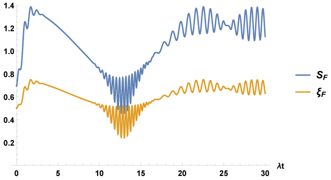


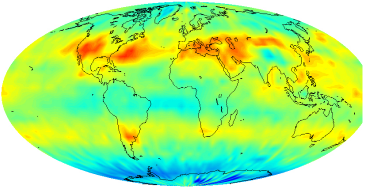


(b) Graph Signal Processing

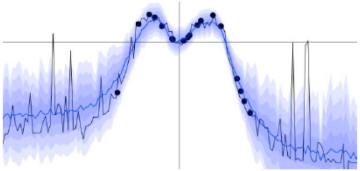


(c) Density of States

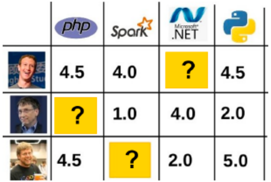




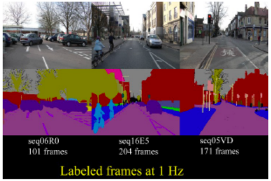
Matrix functions have been utilized in many machine learning problems:



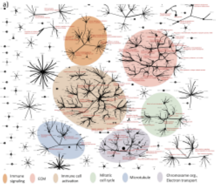
(a) Regression with Gaussian process



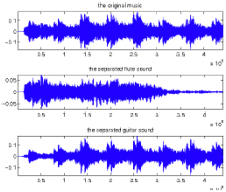
(b) Collaborative filtering for recommendation



(c) Image processing



(d) Gene expression



(e) Speech recognition

Can we estimate $\mathrm{Tr}(f(A))$ faster than matrix multiplication cost?

Discuss *fast scalable* methods with *theoretical guarantees* and perform *well in practice*.

Combine randomization with approximation theory!

Stochastic Chebyshev Method

Chebyshev polynomial approximation

Given a function $f : [-1, 1] \to \mathbb{R}$, a q degree **Chebyshev polynomial** approximation is given by:

$$f(x) \approx p_q(x) = \sum_{j=0}^{q} c_j T_j(x),$$

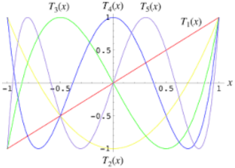
where $T_j(x)$ is the jth degree Chebyshev polynomial with $T_0(x) = 1, T_1(x) = x$,

$$T_{j+1}(x) = 2xT_j(x) - T_{j-1}(x),$$

and the (interpolation) coefficients,

$$c_j = \frac{2 - \delta_{j0}}{\pi} \int_{-1}^{1} \frac{f(x)T_j(x)}{\sqrt{1 - x^2}} dx \quad \text{or} \quad c_j = \frac{2 - \delta_{j0}}{q + 1} \sum_{k=0}^{q} f(x_k)T_j(x_k),$$

with Chebyshev nodes $x_k = \cos\left(\frac{\pi(k+1)/2}{q+1}\right)$.



Stochastic Chebyshev method

- $\boldsymbol{A}$ has spectrum in $[\lambda_{\min}, \lambda_{\max}]$, $\tilde{\boldsymbol{A}} = \left(\frac{2\boldsymbol{A} - (\lambda_{\max} + \lambda_{\min})I}{\lambda_{\max} - \lambda_{\min}}\right)$ spectrum in $[-1, 1]$.
- Approximate $\boldsymbol{x}_l^T f(\boldsymbol{A}) \boldsymbol{x}_l \approx \boldsymbol{x}_l^T \boldsymbol{B} \boldsymbol{x}_l$, where $\boldsymbol{B} = \sum_{j=0}^{q} \tilde{c}_j T_j(\tilde{\boldsymbol{A}})$.
- Let $\boldsymbol{w}_l^{(j)} = T_j(\tilde{\boldsymbol{A}})\boldsymbol{x}_l$; with $\boldsymbol{w}_l^{(0)} = \boldsymbol{x}_l, \boldsymbol{w}_l^{(1)} = \tilde{\boldsymbol{A}}\boldsymbol{x}_l$, and

$$\boldsymbol{w}_l^{(j+1)} = 2\tilde{\boldsymbol{A}}\boldsymbol{w}_l^{(j)} - \boldsymbol{w}_l^{(j-1)}.$$

Stochastic Chebyshev method

- $\boldsymbol{A}$ has spectrum in $[\lambda_{\min}, \lambda_{\max}]$, $\tilde{\boldsymbol{A}} = \left(\frac{2\boldsymbol{A} - (\lambda_{\max} + \lambda_{\min})I}{\lambda_{\max} - \lambda_{\min}}\right)$ spectrum in $[-1, 1]$.
- Approximate $\boldsymbol{x}_l^T f(\boldsymbol{A})\boldsymbol{x}_l \approx \boldsymbol{x}_l^T \boldsymbol{B}\boldsymbol{x}_l$, where $\boldsymbol{B} = \sum_{j=0}^{q} \tilde{c}_j T_j(\tilde{\boldsymbol{A}})$.
- Let $\boldsymbol{w}_l^{(j)} = T_j(\tilde{\boldsymbol{A}})\boldsymbol{x}_l$; with $\boldsymbol{w}_l^{(0)} = \boldsymbol{x}_l, \boldsymbol{w}_l^{(1)} = \tilde{\boldsymbol{A}}\boldsymbol{x}_l$, and

$$\boldsymbol{w}_l^{(j+1)} = 2\tilde{\boldsymbol{A}}\boldsymbol{w}_l^{(j)} - \boldsymbol{w}_l^{(j-1)}.$$

  The **spectral sums** can be estimated as:

$$\mathrm{Tr}(f(A)) \approx \frac{1}{m}\sum_{l=1}^{m}\left[\sum_{j=0}^{q} \tilde{c}_j (v_l)^\top w_l^{(j)}\right]. \tag{1}$$

- **Computational cost**: $O(\mathrm{nnz}(\boldsymbol{A})mq)$.
- *Kernel Polynomial Method* - estimating spectral density [Lin et al.,2016] , Eigencount [Di Napoli et al., 2016], Numerical rank [Ubaru and Saad, 2016, Ubaru et al., 2017].
- *Theoretical analysis* for analytic functions and applications [Han et. al, 2017].

Stochastic Lanczos Quadrature

Stochastic Lanczos Quadrature

- The **Lanczos Quadrature** method by *Gene Golub* and his collaborators in a series of articles.
- Scalar (quadratic form) quantities $\boldsymbol{x}_l^\top f(\boldsymbol{A})\boldsymbol{x}_l$ as *Riemann-Stieltjes integral* problem, and employing *Gauss quadrature rule* to approximate this integral.
- With eigen-decomposition of A as $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^\top$.

$$\boldsymbol{x}_l^\top f(\boldsymbol{A})\boldsymbol{x}_l = \boldsymbol{x}_l^\top \boldsymbol{U} f(\Lambda)\boldsymbol{U}^\top \boldsymbol{x}_l = \sum_{i=1}^{d} f(\lambda_i)\mu_i^2 = \int_a^b f(t)d\mu(t),$$

μ_i are components of $\boldsymbol{U}^\top \boldsymbol{x}_l$ and the measure $\mu(t)$ is a piecewise constant function

$$\mu(t) = \begin{cases} 0, & \text{if } t < a = \lambda_1, \\ \sum_{j=1}^{i} \mu_j^2, & \text{if } \lambda_i \le t < \lambda_{i+1}, \\ \sum_{j=1}^{d} \mu_j^2, & \text{if } b = \lambda_n \le t. \end{cases}$$

- Use Gauss quadrature rule:

$$\int_a^b f(t)d\mu(t) \approx \sum_{k=0}^{q} \omega_k f(\theta_k),$$

$\{\omega_k\}$ are the weights and $\{\theta_k\}$ are the nodes of the q-point Gauss quadrature rule.
- Compute the nodes and the weights via. the *Lanczos algorithm*.

- Use Gauss quadrature rule:

$$\int_a^b f(t)d\mu(t) \approx \sum_{k=0}^{q} \omega_k f(\theta_k),$$

  $\{\omega_k\}$ are the weights and $\{\theta_k\}$ are the nodes of the q-point Gauss quadrature rule.
- Compute the nodes and the weights via. the *Lanczos algorithm*.

- For $A \in \mathbb{R}^{d\times d}$ and $\boldsymbol{x}_l : \|\boldsymbol{x}_l\| = 1$, Lanczos algorithm forms $\boldsymbol{Z}_q^{(l)}$ orthonormal basis for *Krylov subspace*: $\text{Span}\{\boldsymbol{x}_l, \boldsymbol{A}\boldsymbol{x}_l, \dots, \boldsymbol{A}^q\boldsymbol{x}_l\}$,
  and tridiagonal matrix $\boldsymbol{T}_q^{(l)} = \boldsymbol{Z}_q^{(l)\top}\boldsymbol{A}\boldsymbol{Z}_q^{(l)}$.
- The columns $\boldsymbol{z}_j$ of $\boldsymbol{Z}_q^{(l)}$ are related as

$$\boldsymbol{z}_j = p_j(\boldsymbol{A})\boldsymbol{x}_0,\ j = 1, \dots, q,$$

  where $p_j(\cdot)$ are the Lanczos polynomials.
- These polynomials are orthogonal wrt. the measure $\mu(t)$; see Thm 4.2 in [Meurant, Golub, 2009].

Stochastic Lanczos Quadrature

- We approximate,

$$\boldsymbol{x}_l^\top f(\boldsymbol{A})\boldsymbol{x}_l \approx \sum_{k=0}^{q} (\tau_k^{(l)})^2 f(\theta_k^{(l)}) \quad \text{with} \quad (\tau_k^{(l)})^2 = \left[e_1^\top y_k^{(l)}\right]^2,$$

$(\theta_k^{(l)}, y_k^{(l)}), k = 0, 1, ..., q$ are eigenpairs of $\boldsymbol{T}_q^{(l)}$ corresponding to initial vectors $\boldsymbol{x}_l, l = 1, \ldots, m$.

- Matrix function trace estimation as,

$$\mathrm{Tr}(f(\boldsymbol{A})) \approx \frac{n}{m} \sum_{l=1}^{m} \left(\sum_{k=0}^{q} (\tau_k^{(l)})^2 f(\theta_k^{(l)}) \right). \tag{2}$$

- **Computational Cost:** $O(\mathrm{nnz}(\boldsymbol{A})mq)$.

Error Analysis

Theorem

Given a PSD matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ *with its eigenvalues in* $[\lambda_{\min}, \lambda_{\max}]$ *and condition number* $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$, *a function* f *that is analytic inside this interval, and constants* $\epsilon, \eta \in (0,1)$, *for SLQ parameters:*

- $q \geq \frac{\sqrt{\kappa}}{4} \log \frac{K}{\epsilon}$ *number of Lanczos steps, and*
- $m \geq \frac{24}{\epsilon^2} \log(2/\eta)$ *number of starting vectors,*

where $K = \frac{3\lambda_{\max}\sqrt{\kappa}M_\rho}{2m_\rho}$ *with* M_ρ *and* m_ρ *being the absolute maximum and minimum of the function in the interval,*

$$\Pr\Big[\, |\mathrm{Tr}(f(\boldsymbol{A})) - \Gamma| \leq \epsilon \, |\mathrm{Tr}(f(\boldsymbol{A}))| \,\Big] \geq 1 - \eta, \tag{3}$$

where Γ *is the output of the Stochastic Lanczos Quadrature method.*

S. Ubaru, Jie Chen, and Yousef Saad. SIAM Journal on Matrix Analysis and Applications, 38(4), 1075–1099, 2017.

Applications

Further Reading:

- *Applications of trace estimation techniques* by S. Ubaru and Y. Saad.
- *Approximating spectral sums of large-scale matrices using stochastic Chebyshev approximations* by s Insu Han, Dmitry Malioutov, Haim Avron, Jinwoo Shin.
- *Fast Estimation of* $\mathrm{Tr}(f(A))$ *via Stochastic Lanczos Quadrature*, by S. Ubaru, Jie Chen, and Yousef Saad.